Chapter 1
Social Networks Analysis: Tools, Measures and Visualization

Neveen Ghali, Mrutyunjaya Panda, Aboul Ella Hassanien, Ajith Abraham, and Vaclav Snasel

Abstract Social Network Analysis (SNA) is becoming an important tool for investigators, but all the necessary information is often available in a distributed environment. Currently there is no information system that helps managers and team leaders monitor the status of a social network. This chapter presents an overview of the basic concepts of social networks in data analysis including social network analysis metrics and performances. Different problems in social networks are discussed such as uncertainty, missing data and finding the shortest path in a social network. Community structure, detection and visualization in social network analysis is also illustrated. This chapter bridges the gap among the users by combining social network analysis methods and information visualization technology to help a user visually identify the occurrence of a possible relationship amongst the members in a social network. The chapter illustrates an online visualization method for a DBLP

N. Ghali (✉)
Faculty of Science, Al-Azhar University, Cairo, Egypt
e-mail: nev_ghali@yahoo.com

M. Panda
Department of ECE, Gandhi Institute for Technological Advancement (GITA), Bhubaneswar, Odisha, India
e-mail: mrutyunjaya74@gmail.com

A.E. Hassanien
Faculty of Computers and Information, Cairo University, Cairo, Egypt
e-mail: aboitcairo@gmail.com

A. Abraham
Machine Intelligence Research Labs (MIR Labs), Scientific Network for Innovation and Research Excellence, Auburn, WA, USA
e-mail: ajith.abraham@ieee.org

V. Snasel
Faculty of Electrical Engineering and Computer Science, VSB – Technical University of Ostrava, Ostrava – Poruba, Czech Republic
e-mail: vaclav.snasel@vsb.cz

A. Abraham (ed.), *Computational Social Networks: Mining and Visualization*,
DOI 10.1007/978-1-4471-4054-2_1,

(Digital Bibliography & Library Project) dataset of publications from the field of computer science, which is focused on the co-authorship relationship based on the intensity and topic of joint publications. Challenges to be addressed and future directions of research are presented and an extensive bibliography is also included.

Introduction

Social media brings people together in many creative ways, for example users are playing, tagging, working, and socializing online, demonstrating new forms of collaboration and communication that were hardly imaginable just a short time ago. Moreover, social networks play a crucial role in the entrepreneurial process and also help reshape business models and emotions, and open up numerous possibilities to study human interaction and collective behaviour on an unparalleled scale [4, 22, 48, 49].

Nowadays, the Internet plays an increasingly important role and it has gradually infiltrated into every aspect of our lives because of its rich and varied resources. More and more people would like to spend their time on the Internet especially in order to build some kind of large social entertainment community and then try to communicate with each other as frequently as practicable to enable the relationship between them to become closer. Hence, Social Network Analysis (SNA) has become a widely applied method in research and business for inquiring into the web of relationships on the individual, organizational and societal level. With ready access to computing power, the popularity of social networking websites such as Facebook, Twitter, Netlog etc., and automated data collection techniques, the demand for solid expertise in SNA has recently exploded. This interdisciplinary subject is presented herein and introduces the readers to the basic concepts and analysis techniques in SNA to help them understand how to identify key individuals and groups in social systems, to detect and generate fundamental network structures, and then finally to design model growth and diffusion processes in networks. After this introduction to SNA, the readers will be able to design and execute network analysis projects including collecting data and considering ethical and legal implications, to perform systematic and informed analyses of network data for personal, commercial and scholarly use, and to critically review SNA projects conducted by others. It may be concluded that the social network approach for the study of behaviour involves two main themes: (a) the use of formal theory organized in mathematical terms, (b) followed by the systematic analysis of empirical data. The study of social networks really began to take off as an interdisciplinary specialty only after 1970, when modern discrete combinatorics (particularly graph theory) experienced rapid development and relatively powerful computers became readily available. Since then, it has found important applications in organizational behaviour, interorganizational relations, the spread of contagious diseases, mental health, social support, the diffusion of information and animal social organization [44]. Choudhury and Pentland [6] explained how sensors proliferate and how an increasing volume of data is generated from the sensors that contain information about social groupings among sensed individuals. Understanding these groups can be of much importance for a variety of reasons. For example, the DARPA CALO (Cognitive Assistant that Learns and Observes) project [30] exploits knowledge of social groups to anticipate user needs and scheduling conflicts. Physical security and safety applications can then use the knowledge obtained about the social groups to identify unusual, dangerous or threatening behaviour. The MERL (Mitsubishi Electric Research Lab) dataset [54] is an excellent example of the result of daily use of large numbers of sensors in an office environment. This is a dataset of activations of motion sensors distributed through the MERL environment, recorded during the course of 1 year. The sensors generate a high volume of data each day, resulting in a dataset that is difficult to process exhaustively in a short amount of time. The situation is further complicated when one wishes to examine n-way social networks inherent in such a dataset. Jenson and Neville [22] used the Internet Movie Database (IMDb) (www.imdb.com) and the Hollywood Stock Exchange (HSX) (www.hsx.com), an artificial market where players trade in stocks that track the relative popularity of movie actors, both of which are publicly available datasets for research and other non-commercial purposes. The dataset consists of over 300,000 movies, 650,000 persons, and 11,000 studios and the objects are connected by over 2.3 million acted-in links, 300,000 directed links, and 200,000 produced links. The available data on movies vary widely, as not all movies have releases, and HSX data are only available for a small percentage of actors in IMDb. However, the data is more complete for more recent movies and persons.

This chapter is organized in the following way. Section “Social Network Analysis: Basics” provides an explanation of some basic related concepts including social networks versus computer networks and describes some of the social network sites (SNSs). Section “Social Networks: Analysis Metrics and Performance” briefly describes the different performance measures that are encountered during any network analysis. Section “Different Problems in Social Networks” discusses different problems in social networks including uncertainty, missing data in the social network and finding the shortest path. Section “Community Structure and Detection” discuses community structure and detection. Section “Social Network Visualization” discusses visualization in the social networks and briefly illustrates an online analysis tool called FORCOA.net, which is built over the DBLP dataset of scholarly publications in the field of computer science. Finally, conclusions are provided in section “Social Network Concepts to Visualize Terrorist Networks”.

Social Network Analysis: Basics

Social Network Versus Computer Networks

Networks can be categorized according to the topology, which is the geometric arrangement of a computer system. Common topologies include a bus, star, and ring, protocol which defines a common set of rules and signals that computers on the network follow. Network architectures can be broadly classified as either a peer-to-peer or client/server architecture. Computers on a network are sometimes called nodes. Computers and devices that allocate resources for a network are called servers. It is argued that social networks differ from most other types of networks, including technological and biological networks, in two important ways. First, they have non-trivial clustering or network transitivity and second, they show positive correlations between the degrees of adjacent vertices. Social networks are often divided into groups or communities, and it has recently been suggested that this division could account for the observed clustering. Furthermore, group structure in networks can also account for degree correlations. Hence, assortative mixing in such networks with a variation in the sizes of the groups provides the predicted level and compares well with that observed in real-world networks.

Social Network Sites

Social network sites are web sites that allow users to register, create their own profile page containing information about themselves (real or virtual), to establish public ‘Friend’ connections with other members and to communicate with other members [4]. Communication typically takes the form of private emails, public comments written on each others’ profile pages, blog or pictures, or instant messaging. SNSs like Facebook and MySpace are amongst the ten most popular web sites in the world. SNSs are very popular in many countries and include Orkut (Brazil), Cyworld (Korea), and Mixi (Japan).

The growth of SNSs seems to have been driven by the youth, with Facebook originating as a college site [4] and MySpace having an average age of 21 for members in early 2008 [48]. However, an increasing proportion of older members are also using these sites. The key motivating factor for using SNSs is sociability, however, this suggests that some types of people may never use social network sites extensively [49]. Moreover, it seems that extraversion is beneficial in SNSs [42] and that female MySpace users seem to be more extraverted and more willing to self-disclose than male users [40], which suggests they may be more effective communicators in this environment.

SNSs are very interesting because they support public conversations between friends and acquaintances. Walther et al. [51] mentioned that SNS profiles are known as venues for identity expression of members and since public comments appear in these profiles, they may also be composed or interpreted from the perspective of identity expression rather than simply performing a purely communicative function. At the same time, such public conversations are interesting because the web now contains millions of informal public messages that researchers can access and analyze. The availability of demographic information about the sender and recipient in their profile pages makes it more interesting and useful but certain ethical issue might arise (unlike standard interview or questionnaire protocols). However, if the data has been placed in the most public place online as found though Google then its use does not constitute any kind of invasion of privacy [31]. An ethical issue only arises if feedback is given to the text authors or if contact is established. The data mining research on MySpace was more commercially oriented rather than for social science goals, but then an IBM study demonstrated how to generate rankings of musicians based upon opinions mined from MySpace comments [14], and a Microsoft team developed a league table system for movies by extracting lists from MySpace profiles, without explicit sentiment analysis [41].

Social Networks: Analysis Metrics and Performance

This section illustrates the different performance measures that are encountered during any network analysis in order to understand the fundamental concepts. The four most important concepts used in network analysis are closeness, network density, centrality, betweenness and centralization. In addition to these, there are four other measures of network performance that include: robustness, efficiency, effectiveness and diversity. The first set of measures concern structure, whereas the second set concern dynamics and thus depend on a theory explaining why certain agents do certain things in order to access the information.

Social Networks Analysis Metrics

Closeness

This refers to the degree with which an individual is nearer to all others in a network either directly or indirectly [20]. Further, it reflects the ability to access information through the "grapevine" of network members. In this way, closeness is considered to be the inverse of the sum of the shortest distance (sometimes called geodesic distance) between each individual and all others available in the network. For a network with n number of nodes, the closeness is represented mathematically as:

$$c_c(n_j) = \frac{n-1}{\sum_{k=i,j=k}^{n} d(n_i, n_j)} \tag{1.1}$$

where $C_c n_k$ defines the standardized closeness centrality of node j and $d(n_i, n_j)$ denotes the geodesic distance between j and k.

Network Density

Network density is a measure of the connectedness in a network. Density is defined as the actual number of ties in a network, expressed as a proportion of the maximum possible number of ties. It is a number that varies between 0 and 1.0. When density is close to 1.0, the network is said to be dense, otherwise it is sparse. When dealing with directed ties, the maximum possible number of pairs is used instead. The problem with the measure of density is that it is sensitive to the number of network nodes; therefore, it cannot be used for comparisons across networks that vary significantly in size [20].

Centrality: Local and Global

The concept of centrality comprises two levels: local and global. A node is said to have local centrality, when it has a higher number of ties with other nodes, otherwise it is referred to as a global centrality. Whereas local centrality considers only direct ties (the ties directly connected to that node), global centrality also considers indirect ties (which are not directly connected to that node). For example, in a network with a "star" structure, in which, all nodes have ties with one central node, local centrality of the central node is equal to 1.0. Whereas local centrality measures are expressed in terms of the number of nodes to which a node is connected, global centrality is expressed in terms of the distances among the various nodes. Two nodes are connected by a path if there is a sequence of distinct ties connecting them, and the length of the path is simply the number of ties that make it up. The shortest distance between two points on the surface of the earth lies along the geodesic that connects them, and, by analogy, the shortest path between any particular pair of nodes in a network is termed a geodesic. A node is globally central if it lies at a short distance from many other nodes. Such a node is said to be "close" to many of the other nodes in the network, sometimes global centrality is also called closeness centrality. Local and global centrality depends mostly on the size of the network, and therefore they cannot be compared when networks differ significantly in size [20].

Betweenness

Betweenness [20] is defined as the extent to which a node lies between other nodes in the network. Here, the connectivity of the node's neighbours is taken into account in order to provide a higher value for nodes which bridge clusters. This metric reflects the number of people who are connecting indirectly through direct links. The betweenness of a node measures the extent to which an agent (represented by a node) can play the part of a broker or gatekeeper with a potential for control over others. Methodologically, betweenness is the most complex of the measures of centrality to calculate and also suffers from the same disadvantages as local and global centrality. The betweenness of the nodes in a network can be defined as:

$$c_b(n_j) = \frac{xx}{\frac{(n-2)(n-1)}{2}} \tag{1.2}$$

$$xx = \sum_{k<i,j=k,j=t} \frac{g_{kt}(n_j)}{g_{kt}} \tag{1.3}$$

where $c_b(n_j)$ denotes the standardized betweenness centrality of node j, $g_{kt}(n_j)$ represents the number of geodesics linking k and I that contain j in between.

Centralization

Centralization is calculated as the ratio between the numbers of links for each node divided by the maximum possible sum of differences [20]. Centralization provides a measure of the extent to which a whole network has a centralized structure. Whereas centralization describes the extent to which this connectedness is organized around particular focal nodes; density describes the general level of connectedness in a network. Centralization and density, therefore, are important complementary pair measures. While a centralized network will have many of its links dispersed around one or a few nodes, the decentralized network is one in which there is little variation between the number of links each node possesses. The general procedure involved in any measure of network centralization is to look at the differences between centrality scores of the most central node and those of all other nodes. Basically, centralization can be graphed in three ways – one for each of the three centrality measures: local, global and betweenness. All three centralization measures vary from 0 to 1.0. Zero corresponds to a network in which all the nodes are connected to all other nodes whereas a value of 1.0 is achieved on all three measures for "star" networks. However, the majority of real networks lie between these two extremes. Methodologically, the choices of one of these three centralization measures depends on which specific structural features the researcher wants to focus upon. For example, while a betweenness-based measure is sensitive to the chaining of nodes, a local centrality-based measure of network centralization seems to be particularly less sensitive to the local dominance of nodes [20]. It is measured as:

$$R = \frac{\sum_{j=1}^{g}\{\max(D_i) - D_i\}}{(g-1)^2} \tag{1.4}$$

where D_i represents the number of actors in the network that are directly linked to the actor j and g is denoted as the total number of actors present in the network.

Social Network Performance

Once the network analysis is completed, the network dynamics predict the performance of the network which can be evaluated as a combination of: (1) the network's robustness to the removal of ties and/or nodes, (2) network efficiency in terms of the distance to traverse from one node to another and its non-redundant size,

(3) effectiveness of the network in terms of information benefits allocated to central nodes and finally, (4) network diversity in terms of the history of each of the nodes [20].

Robustness

Social network analysts have highlighted the importance of network structure with relation to the network's robustness. The robustness can be evaluated based on how it becomes fragmented when an increasing fraction of nodes is removed. Robustness is measured as an estimate of the tendency of individuals in networks to form local groups or clusters of individuals with whom they share similar characteristics, i.e., clustering. For example, if individuals X, Y, and Z are all computer experts and if X knows Y and Y knows Z, then it is highly likely that X knows Z using the so-called chain rule. If the measure of the clustering of individuals is high for a given network, then the robustness of that network increases – within a cluster/group [20].

Efficiency

Network efficiency can be measured by considering the number of nodes that can instantly access a large number of different nodes – sources of knowledge, status, etc., through a relatively small number of ties. These nodes are treated as non-redundant contacts. For example, with two networks of equal size, the one with more non-redundant contacts provides more benefits than the others. Also, it is quite evident that the gain from a new contact redundant with existing contacts will be minimal. However, it is wise to consume time and energy in cultivating a new contact to un-reached people. Hence, social network analysts measure efficiency by the number of non-redundant contacts and the average number of ties an ego has to traverse to reach any alter, this number is referred to as the average path length. The shorter the average path length relative to the size of the network and the lower the number of redundant contacts, the more efficient is the network [20].

Effectiveness

Effectiveness targets the cluster of nodes that can be reached through non-redundant contacts. In contrast, efficiency aims at the reduction of the time and energy spent on redundant contacts. Each cluster of contacts is an independent source of information. One cluster around this non-redundant node, no matter how numerous its members are, is only one source of information, because people connected to one another tend to know about the same things at about the same time. For example, a network is more effective when the information benefit provided by multiple clusters of contacts is broader, providing better assurance that the central node will be informed. Moreover, because non-redundant contacts are only connected through the central node, the central node is assured of being the first to see new opportunities created by needs in one group that could be served by skills in another group [20].

Diversity

While efficiency is about getting a large number of (non-redundant) nodes, a node's diversity, conversely suggests a critical performance point of view where those nodes are diverse in nature, i.e., the history of each individual node within the network is important. It is particularly this aspect that can be explored through case studies, which is a matter of intense discussion among social network analysts. It seems to suggest that social scientists should prefer and use network analysis according to the first strand of thought developed by social network analysts instead of actor-attribute-oriented accounts based on the diversity of each the nodes [20].

Different Problems in Social Networks

Uncertainty in a Social Network

The uncertainty in digital evidence is not being evaluated at present, thus making it difficult to assess the reliability of evidence stored on and transmitted using computer networks [38]. Uncertainty occurs when the actors are confronted with too many interpretations, causing confusion. In an ambiguous situation there is no lack of information, no gap that could be filled with a better scanning of available information, rather there are at least two (and often more) different interpretations of the situation [2]. Many research works tackled the problem that the data collected through automated sensors, anonymized communication data, and self-reporting logging on Internet-scale networks as a proxy for real relationships and interactions causes some uncertainty.

Gutirrez-Muñoz and Kandel [15] introduced a methodology that incorporates into the social interaction activity records of the uncertainty and time sensitiveness of the events through Fuzzy Social Networks Analysis (FSNA). Also, they investigated an approach based on the analysis of current flows in electrical networks for the extraction of primary routes of interaction among key actors in a social network. They proposed that the ability to capture the influence of all nodes involved in a network over a particular path represents a promising avenue for the extraction of characteristics of the social network assuming that uncertainty and time sensitiveness are parameters of the information stored on activity logs that cannot be ignored and must be accounted for. In Yang et al. [55] an adaptive group Fuzzy analytic network process group decision support system under uncertainty is put forth which makes up for some deficiencies in the conventional analytic network process. In the first step fuzzy judgments are used when it is difficult to characterize the uncertainty by point-valued judgments due to partially known information, and a bipartite graph is formulated to model the problem of group decision making under uncertainty. Then, a Fuzzy prioritization method is proposed to derive the local priorities from missing or inconsistent Fuzzy pairwise comparison judgments. As a result of the unlikeliness for all the decision makers to evaluate all elements under uncertainty, an original aggregation method is developed to cope with the situation where some of the local priorities are missing. Finally, an evaluation of petroleum-contaminated site remedial countermeasures using the proposed group fuzzy analytic network process, indicates that the presented group decision support system can effectively handle uncertainty and support group decision making with high level of user satisfaction. Authors in [16] observed that the characteristics of social systems are poorly modelled with crisp attributes. A concrete agent-based system illustrates the analysis of the evolution of values in a society enhanced with fuzzy logic to improve agent models that get closer to reality. This has been explored in five aspects: relationships among agents, some variable attributes that determine agent states, functions of similarity, evolution of agent states, and inheritance. Vindigni and Janssen [50] proposed a new approach to combine survey data with multi-agent simulation models of consumer behaviour to study the diffusion process of organic food consumption. This methodology is based on rough set theory, which is able to translate survey data into behavioural rules. However, the peculiarity of the rough set approach is that the inconsistencies in a data set about consumer behaviour are not aggregated or corrected since lower and upper approximations are computed. Also rough set data analysis provides a suitable link between survey data and multi-agent models since it is designed to extract decision rules from large quantitative and qualitative data sets.

Missing Data in a Social Network

The inherent problem with much of the data is that it is noisy and incomplete, and at the wrong level of fidelity and abstraction for meaningful data analysis. Thus, there is a need for methods that extract and infer "clean" annotated networks from noisy observational network data. This involves inferring missing attribute values (attribute prediction), adding missing links and removing spurious links between the nodes (link prediction), and eliminating duplicate nodes (entity resolution).

Moustafa et al. [32] identified a set of primitives to support the extraction and inference of a network from observational data, and describe a framework that enables a network analyst to easily implement and combine new extraction and analysis techniques, and efficiently apply them to large observation networks. Perez et al. [36] proposed linguistic decision analysis to solve decision making problems involving linguistic information by using ordinal fuzzy linguistic modelling. In such situations, experts are forced to provide incomplete fuzzy linguistic preference relations. So an additive consistency based estimation process of missing values to deal with incomplete Fuzzy Linguistic preference relations was developed.

Finding the Shortest Path

The problem of finding the shortest path is finding the path with minimum distance or cost from a starting node to an ending node. It is one of the most fundamental network optimization problems. The shortest path problem also has a deep connection to the minimum cost flow problem, which is an abstraction for various shipping and distribution problems, the minimum weight perfect matching, and the minimum mean-cycle problem. Computing shortest paths in graphs is one of the most well-studied problems in combinatorial optimization [33, 46]. The ant colony optimization algorithm is a very efficient machine learning technique for finding the shortest path. The ants, during the activity of finding food and bringing it back to the nest, manage not only to explore a vast area, but also to indicate to their peers the location of the food while bringing it back to the nest. Most of the time, they will find the shortest path and adapt to ground changes, hence proving the great efficiency with which they carry out this difficult task. The authors of [29] proposed the SEMANT algorithm based on ant colony optimization. The proposed algorithm finds the shortest path from every querying peer to one or more appropriate answering peers that possess resources for the given query. An unstructured peer-to-peer network is designed that consists of carefully selected constituents of the ant colony system, AntNet, and AntHocNet, which were combined and adapted for the purposes of the application. Lertsuwanakul and Unger [27] applied the ant colony optimization system where a messenger distributes its pheromone, the long-link details, in the surrounding area. The subsequent forwarding decision has more options to move to, select among local neighbours or send to a node that has a long link closer to its target. They introduced a novel approach for routing in a social network. The authors showed that with additional information, the existence of a shortcut in the surrounding area, they were able to find a shorter path than using the greedy algorithm. Perumbuduru and Dhar [35] proposed the AntNet algorithm by using ant colony optimization. The authors in [26] proposed the Open Shortest Path First protocol by using a genetic algorithm. They implemented a genetic algorithm to find the set of optimal routes to send the traffic from source to destination. A genetic algorithm is well suited for routing problems as it explores solution space in multiple directions at once with less chances to attain a local optimum. The proposed algorithm works on an initial population created by another module, assesses fitness, generates a new population using genetic operators and converges after meeting a specified termination condition.

A hybridization between the ant colony optimization algorithm and genetic algorithm was presented by Cauvery et al. [5] for routing in packet switched data networks. The ant algorithm is found to reduce the size of the routing table. The genetic algorithm cannot use the global information of the network. Hence a combination of these two algorithms, which allows the packets to explore the network independently, helps in finding a path between pairs of nodes effectively. White et al. [52] applied the Ant System with Genetic Algorithm (ASGA) system to the problem of path finding in networks, demonstrating by experimentation that the hybrid algorithm exhibits improved performance when compared to the basic ant system. They demonstrated that the ant system can be used to solve hard combinatorial optimization problems as represented by Steiner vertex identification and shortest cycle determination. The authors in [3] proposed a new neural network to solve the shortest path problem for inter-network routing. The proposed solution extends the traditional single-layer recurrent Hopfield architecture introducing a two-layer architecture that automatically guarantees an entire set of constraints held by any valid solution to of the shortest path problem. This solution aims to achieve an increase in succeeded and valid convergence which is one of the main limitations of previous solutions based on neural networks. Additionally, in general, it requires less neurons. Sang and YI [39] applied a Pulse Coupled Neural Network (PCNN) model called Dual Source PCNN (DSPCNN), which can improve the computational efficiency of pulse coupled neural networks for shortest path problems. Deng and Tong [9] proposed a new algorithm by using a particle swarm optimization algorithm with a priority-based encoding scheme based on a fluid neural network to search for the shortest path in stochastic traffic networks.

Community Structure and Detection

Community structure is one of the key properties of complex networks and detecting communities is a problem of considerable interest. Community structure in the context of networks, refers to the occurrence of groups of nodes in a network that are more densely connected internally than with the rest of the network. This inhomogeneity of connections suggests that the network has certain natural divisions within it. Note that communities are often defined in terms of the partition of the set of vertices, that is each node is put into one and only one community. This is a useful simplification but may not be appropriate in many cases [18]. Identifying meaningful community structure in social networks is inherently a hard problem. Extremely large network size or sparse networks compound the difficulty of the task. Moreover, their scalability is limited to at most a few thousand nodes and execution becomes intractable for very large networks [11]. Among many different community detection approaches, there are two main ones [10]: (1) the graph structure of the network which is named the topology-based community detection approach, and (2) the textual information of the network nodes under consideration which is named the topic-based community detection approach. The detection of the community structure is a promising field of research with many open research challenges. Detecting communities is of great importance in many fields including sociology, biology and computer science, disciplines where systems are often represented as graphs. This problem is very hard and has not yet been satisfactorily solved, despite the huge effort of the large interdisciplinary community of scientists that have been working on it over the past few years.

Community detection in social network analysis is usually considered as a single-objective optimization problem, in which different heuristics or approximate algorithms are employed to optimize an objective function that captures the notion of community. Because of the inadequacy of those single-objective solutions, Shi et al. [43] formulated a multi-objective framework for community detection and presented a multi-objective community detection system (called MOCD) for finding efficient solutions under the framework. The system includes two stages: (1) the community detection stage, and (2) the community selection stage. In the first stage, MOCD simultaneously optimizes two conflicting objective functions with an evolutionary algorithm (EA) and returns a set of solutions, which are optimal in terms of optimization objectives. In order to help decision makers in selecting proper community partitions, in the second stage, two selection approaches are proposed to select one recommendation solution from the solution set returned by the first stage. Through extensive experiments on both simulated and real networks, Shi et al. [43] demonstrated that a combination of two negatively correlated objectives under the multi-objective framework usually leads to remarkably better performance compared with either of the original single objectives, including even many popular algorithms.

Social Network Visualization

Visualizing social networks is of immense help for social network researchers in understanding new ways to present and manage data and to effectively convert the data into meaningful information [47]. A number of visualization tools have been proposed for effective visualization of social networks including Pajek [12], NetVis, Krackplot, IKnow, InFlow, Visone, JUNG and Prefuse, to name a few. Another source of online collaboration has also been visualized to better understand interactions that are provided in a discussion form [8]. Visualizing tasks for better collaboration during software development are proposed in [8] to address issues of co-ordination and geographical distribution of developer teams. Visualizing social networks using query interfaces for wikis and blogs [28] are used to provide end-users with more user-friendly alternatives.

FORCOA.net: An Interactive Tool for Exploring the Significance of Authorship Networks in DBLP Data

To illustrate the importance of visualization, we use the FORCOA.net [19] online tool which provides a nice visualization scheme for Computer Science authors who have publications in selected journals and conferences as registered in DBLP [21]. During the analysis of authors involved in the scholarly publication of articles over

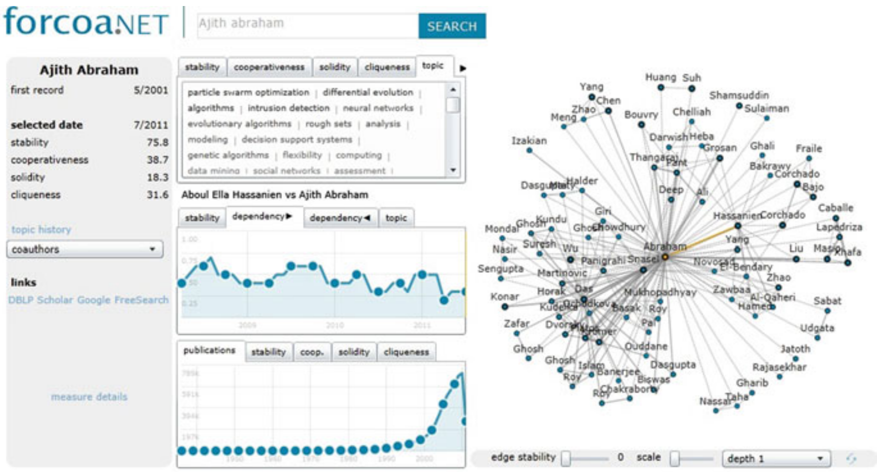


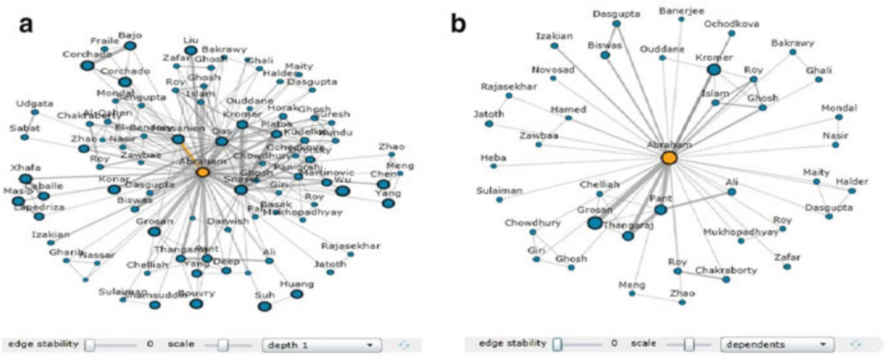
Fig. 1.1 Illustration of Ajith Abraham and Aboul ella Hassanien as a co-author network (01/2012) [19]

a long period of time, the actuality and clarity of the author's collaboration network is often lost. A key requirement was the need for the visualization of an author and his/her academic network in the context of their publication activities. The online tool is built over the DBLP dataset and the dataset contains information about 913,534 authors from the field of computer science and 5,192,020 interactions between these authors [17].

Authors used the stability measure based on the forgetting curve [17, 19]. The stability measure characterizes the behaviour of an author in the network (if the author publishes regularly and over a long term). The online tool is focused on the analysis and visualization of the co-authorship relationship based on the intensity and topic of joint publications. The visualization of co-authorship networks allows one to describe the author and his/her current surroundings, while still incorporating the historical aspects. The analysis is based on using the forgetting function to hold the information relevant to the selected date. Several measures, which can describe different aspects of user behaviour from the scientific social network point of view [17] are also illustrated in the network. In comparison to classical SNA measures (such as centrality, clustering coefficient, etc.) [17, 19] focuses on the usage of edge and vertex stability.

Figure 1.1 illustrates a screenshot of the author Ajith Abraham using the online tool. In the left part of the interface, is a panel containing author details, such as first record in the network and values of several metrics with respect to selected date. This panel also contains a detailed illustration of co-authors and direct links pointing to details about other co-authors [19].

The right part of the interface contains the visualization of the authors social network with current author highlighted. The network can be filtered using some minimum edge weight (see below) or can be switched to a different network view. The default network view contains co-authors to depth 1. The view can be switched



scale = depth-1 scale = dependence

Fig. 1.2 Illustration of Ajith Abraham network with edge stability = 0 and depth = 1 [19]

to depth 2, but also to the so-called dependency or independency network (showing only dependent or independent co-authors).

The middle part contains three groups of panels. The first group contain values of author measures (such as stability, cooperativeness, topics, etc.) over the time period. If you select an edge with a co-author, the second group of panels will appear. These panels contains values of relation measures (stability of the relation, dependency between authors and the topics of the relation). By clicking somewhere in the timeline you can switch the view to a different point in time. The third group of panels contains global values of the whole dataset (such as number of publications, distribution of particular measures over the authors). Figure 1.2 depicts Ajith Abraham's network with a different scale.

Social Network Concepts to Visualize Terrorist Networks

After the 9/11 attacks, much effort was put into developing effective methods for anti-terrorism strategies. Visualization is a very important part of analyzing such a network since it can quickly provide good insight into the network's structure, major members, and their properties [12]. Analyzing huge networks is not an easy task and there is a need to reduce the complexity of these networks, which is usually depicted in the form of huge matrices. The Matrix Factorization Method is a well-established approach and Semi-Discrete Decomposition is highly suitable for dealing with huge networks. Empirical results using the 9–11 network data illustrate the efficiency of the proposed approach [45]. The analysis of general complex networks, link prediction etc. are well illustrated in [1, 7, 13, 23, 24, 34, 37, 53].

The obtained experiment is based on the dataset involving 9/11 attacks from [25]. A binary incidence matrix of involved persons was created and then the rank for Semi-Discrete Decomposition (SDD) factorization was computed and this reduced

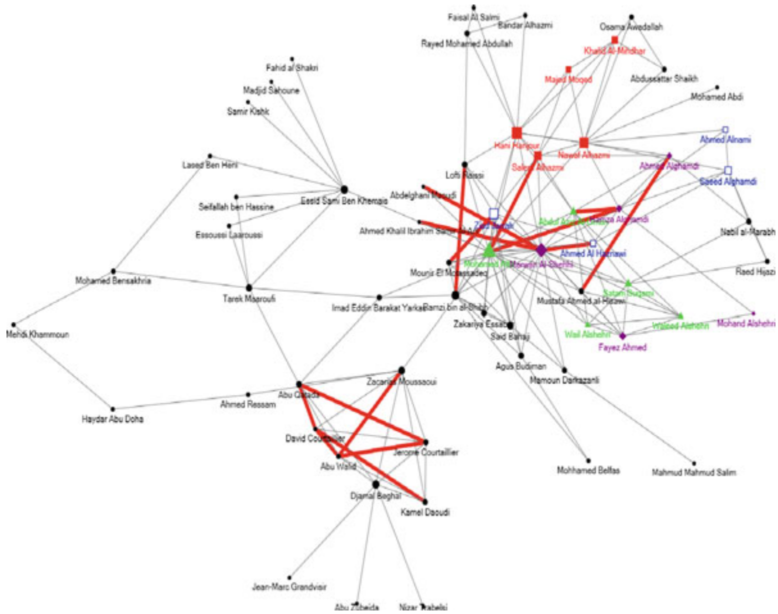


Fig. 1.3 Terrorist network with highlighted link suggestions using SDD reduction (rank 10)

matrix was compared with the original one [45]. A change from zero to one in the reduced matrix can be in a wider sense considered as a link suggestion. In different fields, the suggestion can have different meanings. In the terrorist network, we can consider them as suggestions for investigation to determine whether the link truly exists in reality; for more detail the reader may refer to [45]. The results for a rank parameter setting equal to 10 are illustrated in Fig. 1.3. Same colouring is used as in the original paper [25] by Krebs. Green triangles represent flight AA #11, which crashed into WTC North, full red squares represent flight AA #77 which crashed into the Pentagon, empty blue squares represent flight UA #93 which crashed in Pennsylvania, and full purple diamonds represent flight UA #175 which crashed into WTC South.

Edges drawn in bold red are suggestions obtained by the mentioned reduction. As is evident, the suggested links are in the group of Zacarias Moussaoui, Abu Qatada, David Courtaillier, Jerome Courtaillier, Abu Walid, Kamel Daoudi and Djamal Beghal. This group is also connected using several subgroups in the original data, therefore the proposed method suggests their stronger interconnection. The same holds for the suggested link between Ramzi bin al-Shibh and Lofti Raissi as it connects two different groups of individuals. Remaining suggestions can be explored in a similar way.

Results obtained using a rank parameter setting equal to 20 (that means a lower ratio of reduction) are shown in the right part of Fig. 1.4. Less reduction in this case means less suggestions, but the suggestions obtained for rank 20 are not a subset

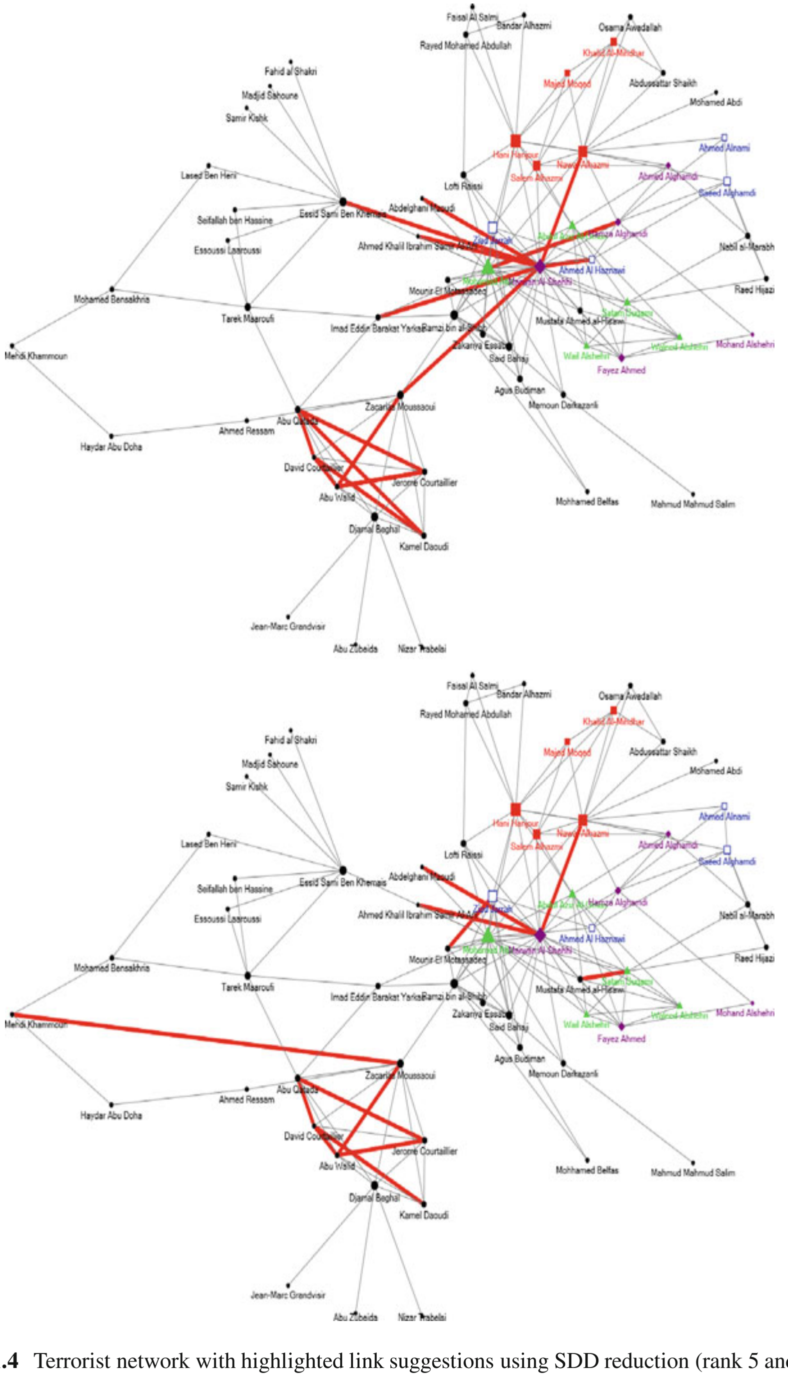


Fig. 1.4 Terrorist network with highlighted link suggestions using SDD reduction (rank 5 and 20)

of suggestions for rank 10. As SDD always tries to minimize the error function, the reduction process is not straightforward – for example the links between Mehdi Khammoun and Zacarias Moussaoui, Mustafa Ahmed al-Hisawi and Satam Suqami as well as the link between Marwan Al-Shehhi and Nawaf Alhazmi are present at rank 20, but disappear at rank 10. The remaining links are still present at rank 10. A similar situation occurs with the setting $k = 5$ (left part of Fig. 1.4), which gives us 16 suggestions – using stronger reduction we have obtained more suggestions, but not all suggestions from rank 10 are present.

Conclusions

The popularity and ease of use of social networking services have excited institutions with their potential in a variety of areas. However, effective use of social networking services poses a number of challenges for institutions including long-term sustainability of the services; user concerns over use of social tools in a work or study context; a variety of technical issues and legal issues such as copyright, privacy, accessibility; etc. Institutions would be advised to consider carefully the implications before promoting significant use of such services. Clear understanding of the structural properties of a criminal network may help analysts target critical network members for removal or surveillance, and locate network vulnerabilities where disruptive actions can be effective. Appropriate network analysis techniques, therefore, are needed to mine criminal networks and gain insight into these problems.

Social Network Analysis (SNA) is becoming an important tool for investigators, but all the necessary information is often distributed over a number of Web servers. Currently there is no information system that helps managers and team leaders to monitor the status of a social network. This Chapter presented an overview of the basic concepts of social networks in data analysis including social network analysis metrics and performances. Different problems in social networks are discussed such as uncertainty, missing data and finding the shortest path in a social network. Community structure, detection and visualization in social network analysis were also discussed.

References

1. Abadie, A., Gardeazabal, J.: Terrorism and the world economy. Eur. Econ. Rev. **52**(1), 1–27 (2008)
2. Antheunis, M.L., Valkenburg, P.M., Peter, J.: Getting acquainted through social network sites: testing a model of online uncertainty reduction and social attraction. Comput. Hum. Behav. **26**, 100–109 (2010)
3. Araujo, F., Ribeiro, B., Rodrigues, L.: A neural network for shortest path computation. IEEE Trans. Neural Netw. **12**(5), 1067–1073 (2001)

4. Boyd, D., Ellison, N.: Social network sites: definition, history and scholarship. J. Comput. Mediat. Commun. **13**(1), 210–230 (2007)
5. Cauvery, N., Viswanatha, K.V.: Routing in dynamic network using ants genetic algorithm. IJCSNS Int. J. Comput. Sci. Netw. Secur. **9**(3), 194–200 (2009)
6. Choudhuri, T., Pentland, A.: Sensing and modelling human networks using the sociometer. In: Proceedings of 7th IEEE Symposium on Wearable Computing, New York (2003)
7. Czinkota, M.R., Knight, G.A., Liesch, P.W., Steen, J.: Positioning terrorism in management and marketing: research propositions. J. Int. Manag. **11**(4), 581–604 (2005)
8. De Nooy, W., AMrvar, A., Batagelig, V.: Exploring SNA with pajek. Cambridge University Press, Cambridge/New York (2004)
9. Deng, Y., Tong, H.: Dynamic shortest path algorithm in stochastic traffic networks using PSO based on fluid neural network. J. Intell. Learn. Syst. Appl. **3**, 11–16 (2011)
10. Ding, Y.: Community detection: topological vs. topical. J. Informetr. **5**(4), 498–514 (2011)
11. Fortunato, S.: Community detection in graphs. Phys. Rep. **486**(3–5), 75–174 (2010)
12. Freeman, L.: Visualizing social network. J. Soc. Struct. **1**(1) (2000)
13. Frey, B.S.: How can business cope with terrorism? J. Policy Model. **31**(5), 779–787 (2009)
14. Grace, J., Gruhl, D., et al.: Artist ranking through analysis of on-line community comments. IBM Research technical report (2007)
15. Gutirrez-Muñoz, A., Kandel, A.: Current flows in electrical networks for fuzzy social network analysis (FSNA). Department of Computer Science and Engineering, University of South Florida (2009)
16. Hassan, S., Garmendia, L., Pavon, J.: Introducing uncertainty into social simulation: using fuzzy logic for agent-based modelling. Int. J. Reason. Based Intell. Syst. **2**(2), 118–124 (2010)
17. Horak, Z., Kudelka, M., Snasel, V., Abraham, A., Rezankova, H.: Forcoa.NET: an interactive tool for exploring the significance of authorship networks in DBLP Data. In: Third International Conference on Computational Aspects of Social Networks (CASoN 2011), Salamanca, pp. 261–266. IEEE (2011). ISBN: 978-1-4577-1131-2
18. http://en.wikipedia.org/wiki/Community_structure. Accessed on March 2012
19. http://www.FORCOA.net. Accessed on 07 Jan 2012
20. http://en.wikipedia.org/wiki/Betweenness. Accessed on 07 Jan 2012
21. http://www.informatik.uni-trier.de/ley/db/. Accessed on 07 Jan 2012
22. Jenson, D., Neville, J.: Data mining in networks. In: Symposium on Dynamic Social Network Modelling and Analysis, National Academy of Sciences. National Academy Press, Washington (2002)
23. Koh, W.T.H.: Terrorism and its impact on economic growth and technological innovation. Technol. Forecast. Soc. Change **74**(2), 129–138 (2007)
24. Kollias, C., Messis, P., Mylonidis, N., Paleologou, S.: Terrorism and the effectiveness of security spending in Greece: policy implications of some empirical findings. J. Policy Model. **31**(5), 788–802 (2009)
25. Krebs, V.E.: Uncloaking terrorist networks. First Monday **7** (2002)
26. Kumar, R., Kumar, M.: Exploring genetic algorithm for shortest path optimization in data networks. Glob. J. Comput. Sci. Technol. (GJCST 2010) **10**(11), 8–12 (2010)
27. Lertsuwanakul, L., Unger, H.: An improved greedy routing algorithm for grid using pheromone-based landmark. World Acad. Sci. Eng. Technol. **59**, 172–176 (2009)
28. Matsuo, Y., et al.: Polyphonet: an advanced social network extraction system from the web. In: www 06, Proceedings of International Conference on World Wide Web, New York, pp. 397–406 (2006)
29. Michlmayr, E.: Ant algorithms for self-organization in social networks. Ph.D. thesis, Vienna University of Technology, Faculty of Informatics (2007)
30. Mitchell, T., wang, S., Huang, Y., Cheyer, A.: Extracting knowledge about users activities from raw workstation contents. In: Proceedings of 21st National confefernce on Artificial Intelligence, AAAI-2006, Boston (2006)
31. Moor, J.H.: Towards a theory of privacy for the information age. In: Spinello, R.A., Tavani, H.T. (eds.) Readings in CyberEthics, 2nd edn., pp. 407–417. Sudbury, MA: Jones and Bartlett (2004)

32. Moustafa, W., Deshpande, A., Namata, G., Getoor, L.: Declarative analysis of noisy information networks. In: Proceedings of IEEE 27th International Conference on Department of Computer Science, Data Engineering Workshops (ICDEW), Hannover, pp. 106–111 (2011)
33. Mukhef, H.A., Farhan, E.M., Jassim, M.R.: Generalized shortest path problem in uncertain environment based on PSO. J. Comput. Sci. **4**(4), 349–352 (2008)
34. Paraskevas, A., Arendell, B.: A strategic framework for terrorism prevention and mitigation in tourism destinations. Tour. Manag. **28**(6), 1560–1573 (2007)
35. Perumbuduru, S., Dhar, J.: Performance evaluation of different network topologies based on ant colony optimization. Int. J. Wirel. Mob. Netw. (IJWMN) **2**(4), 141–157 (2010)
36. Prez, I.J., Alonso, S., Cabrerizo, F.J., Herrera-Viedma, E.: A fuzzy linguistic decision support system to aid users in e-commerce activities. In: Proceedings of the 2010 World Congress in Computer Science, Computer Engineering, and Applied Computing (WORLDCOMP 2010). The 2010 International Conference on Artificial Intelligence (ICAI 2010), Las Vegas, 12–15 July 2010
37. Reid, E.F., Chen, H.: Mapping the contemporary terrorism research domain. Int. J. Hum. Comput. Stud. **65**(1), 42–56 (2007)
38. Saint-Charles, J., Mongeau, P.: Different relationships for coping with ambiguity and uncertainty in organizations. Soc. Netw. **31**, 33–39 (2009)
39. Sang, Y., YI, Z.: A modified pulse coupled neural network for shortest path computation. J. Comput. Inf. Syst. **6**(9), 3095–3102 (2010)
40. Schrock, A., Examining social media usage: technology clusters and social network relationships. First Monday **14**(1–5), January (2009)
41. Shani, G., Chickering, M., Meek, C.: Mining recommendations from the web. In: Proceedings of 2008 ACM Conference on Recommender System, Lausanne, pp. 35–42 (2008)
42. Sheldon, P.: The relationship between unwillingness to communicate and students Facebook use. J. Media Psychol. Theor. Method Appl. **20**(2), 67–75 (2008)
43. Shi, C., Yan, Z., Cai, Y., Wu, B.: Multi-objective community detection in complex networks. Appl. Soft Comput. **12**(2), 850–859 (2012)
44. Snasel, V., Harak, Z., Abraham, A.: Understanding social networks using formal concept analysis. In: Web Intelligence and Intelligent Agent Technology, 2008, WI-IAT '08, Sydney, 390–393 (2008)
45. Snasel, V., Horak, Z., Abraham, A.: Link suggestions in terrorists networks using semi discrete decomposition. In: Sixth International Conference on Information Assurance and Security (IAS), pp. 337–339. IEEE (2010). ISBN 978-1-4244-7408-0
46. Sommer, C.: Approximate shortest path and distance queries in network. Ph.D. thesis, Department of Computer Science Graduate School of Information Science and Technology, The University of Tokyo (2010)
47. Tantipathananandh, C., Breger-wolf, T., Kempe, D.: A framework for community identification in dynamic social network, In: Proceedings of KDD 2007, pp. 717–726. San Jose, California, USA (2007)
48. Thellwal, M.: Social networks, gender and friending, analysis of Myspace profiles. J. Am. Soc. Inf. Sci. Technol. **591**(8), 1321–1330 (2008)
49. Tufekci, Z.: Groming, gossip Facebook and Myspace: what can we learn about these sites from those who wont assimilate? Inf. Commun. Soc. **11**(4), 544–564 (2008)
50. Vindigni, G., Janssen, M.A.: Organic food consumption a multi-theoretical framework of consumer decision making, The current issue and full text archive of this journal is available at http://www.emeraldinsight.com/0007070X.hm. Accessed on 20 Mar 2012 (2002)
51. Walther, J., Ander Heide, B., Kim, S. Westerman, D., Tang, S.T.: The role of friends appearance and behaviour on evaluations of individuals on facebook: are we known by the company we keep? Hum. Commun. Res. **34**, 28–49 (2008)
52. White, T., Pagurek, B., Oppacher, F.: ASGA: improving the ant system by integration with genetic algorithms. In: Proceedings of the Third Annual Conference, University of Wisconsin, Madison, pp. 610–617 (1999)

53. Wolf, Y., Frankel, O.: Terrorism: toward an overarched account and prevention with a special reference to pendulum interplay between both parties. Aggress. Violent Behav. **12**(3), 259–279 (2007)
54. Wren, C.R., Ivanov, Y.A., Leign, D., Westhues, J.: The MERL motion detector datasets: 2007 Workshop on Massive Datasets. Mitsubishi electric research laboratories tech report, TR-2007-069
55. Yang, W.Z., Ge, Y.H., He, J.J., Liu, B.: Designing a group decision support system under uncertainty using group Fuzzy analytic network process (ANP). Afr. J. Bus. Manag. **4**(12), 2571–2585 (2010)